The Prisoners’ Dilemma and the Problem of Cooperation

One of the central problems of international politics is the problem of cooperation. How can governments reach agreements that make them better off? Arms races provide a good example of the problem of cooperation. On the one hand, if all governments spend a lot of money on military forces, no government gains any additional security or is better able to influence others. Thus, countries are better off if they can somehow all agree to refrain from engaging in a military buildup—they will enjoy the same level of security without an arms race as they do with an arms race, but without an arms race each country saves the resources it otherwise dedicates to the military. In other words, there are gains to be had from international cooperation to limit military expenditures. Achieving these gains is difficult, however, due to the structure of the international system. This problem of cooperation is more than a theoretical problem—throughout the Cold War the United States and the Soviet Union tried, with varying degrees of success, to cooperatively manage their nuclear competition. The purpose of this short reading is to use game theory to demonstrate why international cooperation is so difficult.

The problem of cooperation in general, and arms races in particular, can be modeled with game theory. Game theory is an approach to the study of interdependent decision-making, often called strategic interaction, developed by mathematicians and economists. One game in particular, the prisoners= dilemma, has received the most attention as a model of how strategic interaction in the anarchic international system creates incentives for governments to enter into arms races and complicates their abilities to effectively end arms races.

In the prisoners= dilemma, two governments, lets call them the United States and the Soviet Union must decide whether to build nuclear weapons or not to build nuclear weapons. In the terminology of game theory, we say that each government has two strategy choices: build nuclear weapons, which we will denote as ***b***, not build nuclear weapons, which we will denote as ***n***. Two governments with two strategy choices each generates the two-by-two matrix depicted in figure 1. Memorize this matrix because it is critically important. Each cell in this matrix corresponds to a combination of American and Soviet strategies, and these strategy combinations produce real-world outcomes.

Figure 1: The Prisoners’ Dilemma and Arms Races

		United States	
		Not Build	Build
Soviet Union	Not Build	***n,n*** ***(3,3)***	***n,b*** ***(1,4)***
	Build	***b,n*** ***(4,1)***	***b,b*** ***(2,2)***

Preference Orders:
Soviet Union: ***bn*** > ***nn*** > ***bb*** > ***nb***
United States: ***nb*** > ***nn*** > ***bb*** > ***bn***

We can describe these outcomes starting in the top left cell and moving clockwise. It is important to say a word about the notation we will use before we proceed. By convention we list the row player=s (the player who selects its strategy from the rows of the matrix) strategy choice first and the column player=s (the player who selects its strategy from the columns of the matrix) strategy choice second. Thus, the strategy combination referred to as ***"bn@*** means that the row player (the Soviet Union) has chosen the strategy "build nuclear weapons" and the column player (the United States) has chosen the strategy "do not build nuclear weapons." We can now describe the four outcomes in the prisoners= dilemma. If the US chooses do not build and the Soviet Union chooses "do not build" (***nn***), then the two governments are not engaged in an arms race. If the Soviet Union chooses "do not build" and the US chooses "build nuclear weapons" (***nb***), then the US gains a power advantage over the Soviet Union. If the Soviet Union chooses "build nuclear weapons" and the US chooses "build nuclear weapons" (***bb***), then the two countries are engaged in an arms race. Finally, if the Soviet Union chooses "build nuclear weapons" and the US chooses "do not build" (***bn***), then the Soviet Union gains a power advantage over the US.

Now we must determine how each government ranks these four outcomes: which is their most, second most, third most, and least preferred outcome? The Soviet Union's most preferred outcome is ***bn***, where it builds nuclear weapons and the US does not because in this outcome the Soviet Union gains power relative to the United States. Soviet security is thereby enhanced. The Soviet Union's least preferred outcome is ***nb*** because in this outcome the US gains power relative to the Soviet Union. Soviet security is thereby diminished. We have the most preferred and the least preferred outcomes. Where do the other two outcomes fit in? The Soviet Union prefers ***nn*** to ***bb*** because if both governments build nuclear weapons or if both governments do not build nuclear weapons, their relative power remains constant. Yet, if both build nuclear weapons, each spends money on nuclear weapons that could have been used for other purposes. Thus, ***nn*** is better than ***bb*** because the Soviet Union saves money. It should also be clear that the Soviet Union prefers the outcomes ***nn*** and ***bb*** less than ***bn***, because with ***bn*** the Soviet Union gains a relative power advantage over the US. Finally, the Soviet Union prefers ***nn*** and ***bb*** more than ***nb***, because under ***nb*** the Soviet Union suffers a relative power loss. Thus, we have a clear preference order for the Soviet Union ***bn*** > ***nn*** > ***bb*** > ***nb*** where the "greater than" sign means "is preferred to@.

What about the US? The prisoners= dilemma is a symmetric game, which means that the US faces the exact same situation as the Soviet Union. Because the US faces a situation identical to the Soviet Union, its payoff order will be identical to the Soviet Union's with one small difference arising from the notation we use. Like the Soviet Union, the US=s most preferred outcome is the one in which it gains a relative power advantage, but for the US this is the outcome ***nb***. And also like the Soviet Union, the US's least preferred outcome is the one in which the Soviet Union gains a power advantage, but for the US this is the outcome ***bn***. Thus, the US payoff order is identical to the Soviet Union's payoff order, but the position of the most and least preferred outcomes are reversed: ***nb*** > ***nn*** > ***bb*** > ***bn***.

How will the United States and the Soviet Union play this game, that is, what strategies will they select, and what outcome should we expect? In the prisoners= dilemma each actor has what is called a ***dominant strategy***. We can make this clear by working through the Soviet Union's best responses to US strategy choices. If the US plays the strategy "do not build nuclear weapons," the Soviet Union has to choose between building nuclear weapons and not building nuclear weapons. If the Soviet Union opts for "do not build" in response to the US decision to build, the Soviet Union receives its second most preferred outcome. If the Soviet Union decides to build nuclear weapons in response to the US play of "do not build," the Soviet Union receives its most preferred outcome. Thus, if the US plays do not build, the Soviet Union's best response is to build nuclear weapons. Now suppose that the US decides to build nuclear weapons. If the Soviet Union plays "do not build" in response to the US decision to build nuclear weapons, the Soviet Union receives its least preferred outcome. If the Soviet Union decides to build nuclear weapons in response to the US decision to build nuclear weapons, the Soviet Union gets its second least-preferred outcome. Thus, if the US builds nuclear weapons, the Soviet Union's best response is to build nuclear weapons. In the prisoners= dilemma, therefore, the choice "build nuclear weapons" provides the Soviet Union with a higher payoff than "do not build" regardless of the strategy the US plays. Therefore, the strategy "build nuclear weapons" is said to "dominate" the strategy "do not build" in the prisoners' dilemma. Building nuclear weapons is always preferred to not building nuclear weapons.

Because the prisoners= dilemma is symmetric, build nuclear weapons is also the US=s dominant strategy. Because both governments have dominant strategies to build nuclear weapons, the game always yields the same outcome: both governments build nuclear weapons, and the game produces the ***bb*** outcome. In other words, the prisoners dilemma suggests that the United States and the Soviet Union are likely to find themselves engaged in a nuclear arms race that enhances neither country's security,

There are three important things to recognize about the build nuclear weapons-build nuclear weapons outcome in the prisoners= dilemma. First, this outcome is ***Pareto sub-optimal***. Pareto optimality is a way to conceptualize societal welfare. An outcome is Pareto optimal when no single individual can be made better off without at the same time making another individual worse off. Pareto sub-optimal, therefore, refers to outcomes in which it is possible for at least one individual to realize a welfare improvement without making any one else in that society worse off. In the prisoners= dilemma the build nuclear weapons-build nuclear weapons outcome is Pareto sub-optimal because both governments are better off with the outcome ***nn*** than they are with the outcome ***bb***. Thus, rational behavior on the part of each individual government, that is each playing their dominant strategy "build nuclear weapons", produces a sub-optimal collective outcome: the United States and the Soviet Union engage in a nuclear arms race even though both would be better off if both played the "do not build" strategy.

Second, the ***bb*** outcome is a ***Nash equilibrium***. A Nash equilibrium is an outcome at which neither player has an incentive to change strategies unilaterally. Once the two governments arrive at the build nuclear weapons-build nuclear weapons outcome neither has an incentive to change its strategy unilaterally. If the Soviet Union changes its strategy from build nuclear weapons to do not build the outcome shifts to ***nb***, the Soviet Union's least preferred outcome. Thus, the Soviet Union has no incentive to change strategies unilaterally. If the US changes its strategy from build nuclear weapons to do not build the outcome moves to ***bn***, the US's least preferred outcome. Thus, the US has no incentive to change strategies unilaterally. Because neither the Soviet Union nor the US has an incentive to change strategies unilaterally once they arrive at ***bb***, the build nuclear weapons-build nuclear weapons outcome is a Nash equilibrium. Putting these first two points together, the prisoners= dilemma=s central expectation is that governments find themselves stuck in an arms race even though they could all realize gains from an end to this arms race, and neither side will have an incentive to change its behavior to bring this arms race to an end.

This points us to the third important thing to recognize about the prisoners= dilemma. The central factor preventing governments from realizing the gains available from mutual restraint in nuclear weapons programs is the lack of a mechanism with which to enforce agreements. If there existed a third party (the equivalent of a police force and the judiciary in domestic political systems) to enforce an agreement, then it would be possible for the two countries to achieve the cooperative outcome. With an effective enforcement mechanism the US and the Soviet Union could agree to play "do not build" strategies and, because cheating on this agreement would be punished, both would abide by the agreement. In the anarchic international system, however, no third party capable of enforcing agreements exists. Without an enforcement mechanism neither the United States nor the Soviet Union has an incentive to trust the other to abide by any agreement they make. Unwilling to risk facing their least preferred outcome in which they show restraint while the other country increases its nuclear power, both governments will play their dominant strategies. The anarchic nature of the international system, therefore, creates incentives for governments to engage in arms races, and makes it difficult for governments to bring these arms races to an end.

The more general point about international politics highlighted by the prisoners' dilemma is the following: the anarchic structure of the international system makes it difficult for governments to cooperate. Even though the United States and the Soviet Union would both be better off if they could cooperate and limit their nuclear weapons programs, both will continue to engage in arms race because there is no mechanism to ensure that each will comply with any cooperative agreement they reach. Thus, the prisoners' dilemma highlights how the weakness of political institutions in the international system affects the way governments behave in international politics.